

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The findings of the work described in this thesis (Chapters 2, 3 and 4) lead to a number of conclusions which have been summarised here.

The results of the first stage of the work suggest that the assessment has been highly reliable, even when the teachers assessed two skills at the same time, instead of only one. However, further study may be carried out to determine the maximum number of skills that one teacher can assess in one session for a set of approximately 20 pupils and still maintain high standards of reliability.

Although the achievement in pouring out a liquid skill is very good, for the volume measurement skill, it is only about 33% of the total marks. This is perhaps due to the fact that the skill of volume measurement is a more technically oriented task involving the precision of reading a finely graduated scale. Furthermore the relatively poor performance in volume measurement reflects how poor some pupil's ability is in carrying out some manipulative skills as has been mentioned earlier by Bryce et.al.²². No matter how simple a skill may be, it must be taught with great care so that pupils of even low ability range, who need extra help, can acquire the skill.

Lack of significant correlation (shown by Pearson-r) between the achievement data obtained from assessment by Method -T and that from assessment by Method -P suggest that marking of the reported results can not be used as a substitute for the assessment of manipulative skill during the practical session. The lack of statistical significance does not necessarily mean that there is no educational significance of collecting achievement data from different ways of assessment of pupils carrying out

a skill in the laboratory. Some pupils can express their skill and abilities best either in preparing written report or by performing in the laboratory, but not in both ways. So, at least for these pupils it would be fair to assess them in both ways, instead of considering one method of assessment as a substitute of others, no matter how convenient and time saving it could be for the teacher. The low Pearson-r value may prevent the substitution of Method -T by Method -P of assessment, but it may encourage the use of both methods of assessment as a complimentary to each other.

Agreement among the judges (another measure of the reliability of assessment) is almost 100% for the pouring out liquid skill, while it is about 50% for the volume measurement skill. This is probably due to the fact that the pouring skill is simpler than the volume measurement skill, as has been seen from the pupils' achievement in performing these skills. In measuring a certain volume of liquid using a measuring cylinder, the pupil needs to perform more precision work as he/she is expected to bring the lower meniscus of the liquid column at the correct mark on the scale. Regarding the judges' tasks, as they have to follow a more complex check-list for the volume measurement skill assessment, there is room for a variation in their opinion and hence lack of agreement.

Further research using a check-list of various degrees of complexity for different experiments can be carried out to ascertain whether the reliability of assessment changes, and if so, to what extent.

The results of the second stage of the research show that the taught groups can produce better results than the untaught groups in carrying out manipulative skills. From this, it may be suggested that the pupils can be

expected to remember a skill taught even 8 weeks earlier. However, this ability to retain and reproduce a scientific skill is likely to vary with its nature and the complexity involved. For example, the achievement of the taught group in performing the volume measurement skill is double that of the untaught groups, whereas in the heating skill, the taught group has done slightly better than the untaught groups (overall score being very high for all the groups). This difference in the standard of achievement by the taught group in 2 skills is probably due to the fact that the pupils' learning ability depends on the degree of complexity of the skills and their retention ability depends on how this knowledge is anchored in the pupils' minds and assimilated with his/her existing ideas^{37,38}.

The results of the studies of pupils' ability to transfer skills to a new situation (3rd stage of the research) shows that the taught group can produce better results than the untaught groups. As the achievement of the untaught groups is also very high, one may suggest that the cause for this is that the pupils participating in the assessed practical test are of very high academic ability and have talents to perform a new skill without any formal lesson on that (or any similar) skill.

However the higher achievement of the taught groups indicates that teaching does indeed help pupils to improve their knowledge, some of which is retained and enables him/her to tackle a similar problem more efficiently. In future, work can be carried out using pupils of a wide ability range and asking them to execute a series of skills with varying degrees of complexity and difficulty.

Pearson-r values showed no significant correlation between the

achievement data obtained from two modes of assessment (Method -T and Method -P) for the volume measurement skill using a burette in the third stage of this investigation. Lack of such relationship has also been found, mentioned earlier, in the first stage of this investigation. So it would not be fair to assess the skills involved in an experiment only on the basis of the final written report/result, unless it is clearly established which skill (or skills) has predominant influence on the final outcome of the experiment. This could be due to the fact that the outcome of the extended task/pupils' written report is not strongly affected by the skills assessed in the laboratory. One needs to recognise that, apart from the skills assessed, there are a number of skills involved in the experiment which have not been assessed and could have influenced the final result supplied in the written report. So all possible skills/steps involved in the experiment have to be isolated and further studies of this type have to be carried out. If a strong correlation exists between the pupils' achievement data gathered from two sources, (a) assessment by Method -T (on the spot observation of pupils' performance of the manipulative skills/technique) and Method -P (written report of the whole task), then these skills could be assessed by Method -P instead of Method -T and the assessment will be valid for these specific skills only. The results of this study support the view, suggested earlier by Bryce and Robertson²⁶, that it is more important and meaningful to assess manipulative skills by observation on the spot during experimental work, rather than depending solely on the written report containing answers to some questions or the recording of some observations, although the latter method has been advocated by Woolnough and Toh³⁰. In their published work, Woolnough and Toh³⁰ gave detailed information about how a number of reporting methods had

been investigated and concluded that a "broadly cued" method of reporting by the pupil proved to be the best in reflecting pupils' laboratory performance which can be assessed by the teacher at a later stage.

Although they have reported a strong correlation (Pearson $-r$) between pupils' achievement assessed by teacher observing them directly in the laboratory, and that while assessed on the basis of their written report, they have not given information regarding the methodology (organisational aspect and a check list) used for on-the-spot assessment. In this case the strong correlation between two sets of achievement data would have been meaningful and the case for replacing one method of assessment by the other would have been strong, if they had provided information about the nature of the skill assessed by direct observation, and also the check-list used.

The findings of this research are likely to have some implications on further research in this field, and also on classroom activities involving teacher and pupil. One could extend this type of research work for other age groups to gain a more generalised picture of the conclusion drawn here. Practical work carried out by lower age group pupils can be considered as pre-GCSE practice and preparation for the assessed practicals, forming part of the GCSE examination which will be faced by the pupils at the age of 15+. So, it would be beneficial for both teachers and pupils alike to try to become involved in assessment (perhaps of the formative type) of practical work from an early stage of secondary school education. As assessed practical is an integral part of the GCSE examination, practice and preparation in this aspect from an early age will undoubtedly help the pupils to do well in the final examination. Furthermore pupils tend to work with more application if it is known to them that the practical work

is being assessed. Such a practice will help the teachers also. They will use this experience to improve their teaching technique, develop a more effective long-term teaching plan, and produce a reliable assessment scheme. In the end a national assessment scheme may be developed so that a parity in the marking standard is maintained throughout the schools and thus the need for rigorous moderation is reduced.

Further research will help to establish the confidence in the reliability of teachers' assessments made by direct observation of pupils' performance in-situ. For this purpose, one needs to find out (1) if the reliability of assessment varies from teacher to teacher or from school to school, (2) if it varies as a teacher increases the items on the check-list (skills to be assessed) in a session, and (3) the maximum number of skills one teacher can assess in a session without compromising the reliability.

Reduction in the number of pupils per session/set to maintain/improve the reliability of assessment is not a realistic approach because a teacher, being in charge of a set, will be expected to assess all his pupils by himself in a session. Even if he does not assess all pupils, but decides to assess only a small section in every lesson for one skill, he will have to look after them and make sure that the un-assessed group spends their lesson-time in useful academic activities without disturbing the assessed group.

Further research in this field would be needed to establish the reliability and validity of assessment of practical work on the basis of the outcome of the extended task or the final product.

In this thesis, comparison between the performance of boys and that of

girls has not been considered, as an innumerable number of studies has already been done previously in this area. I feel that it may be a more realistic idea not to give too much importance on the gender difference of the pupils, and treat them equally for teaching or assessment purposes.

More research of the type described in this thesis will help to develop a suitable check-list for a well defined mark-scheme of all GCSE practicals to maintain reliability of assessment between various schools, again helping the moderators' task. Without a proper check-list and a well defined mark allocation scheme, teachers may adopt the method of subjective judgement (impression grading) and this may lead to erratic/irrational marking based solely on the opinion of the assessor which may vary markedly. Therefore, well defined check-lists for assessment of manipulative skills in science practicals will improve the reliability of assessment (see Eglen and Kempa⁴⁸). It is also necessary to establish the total number of skills a teacher could be expected to assess single handedly in a class of 20-25 pupils in one practical session (say, a double period) without lowering the standard of assessment (i.e. compromising the reliability of assessment).

Further research on pupils' skill retention and transfer ability can be carried out to obtain more information regarding the cognitive aspect of practical work and make some advancement in the understanding of pupils' learning process during a practically orientated teaching session.

The teaching and learning of the practical skills involved in science are laboratory based work. Assessment based only on written work (written report of pupils' findings from the practical work) will neglect the

manipulative aspects of the practical work which become visible only during the performance of the task itself in the laboratory.

Having carried out further research, one would be able to produce a checklist of manipulative skills for an experiment, which are difficult to perform, subject to improvement by good teaching and thus capable of influencing the final results. If it is established by the researcher that there is a strong correlation between the pupils' achievement data obtained by assessment method -T (assessment of the "techniques", i.e. performance of key manipulative skill on the spot in the laboratory) and that obtained by assessment method -P (assessment of the "product" i.e. outcome of the whole task submitted as a written report), the latter method can be used as a substitute of the former and teachers can assess pupils' practical skill without the constraints of the classroom. However, until such a situation is reached, it would be a wise policy to assess pupils from both the angles. The results obtained from two modes of assessment would be complementary, providing a more complete picture of the pupil's achievement profile concerning their ability to carry out the practical work involved in school science courses.